

Speciation-rate dependence in species-area relationships

Simone Pigolotti ^{a,b} and Massimo Cencini ^c

^a*The Niels Bohr International Academy, The Niels Bohr Institut, Blegdamsvej 17, DK-2100 Copenhagen, Denmark*

^b*Corresponding author, email: pigo@nbi.dk, Ph. +4535325238, Fax +4535325425*

^c*INFN-CNR, SMC Center for Statistical Mechanics and Complexity, Dipartimento di Fisica Università di Roma "La Sapienza" Piazzale A. Moro, 2 00185 Rome, Italy and ISC-CNR, Via dei Taurini, 19 00185 Rome, Italy.*

Abstract

The general tendency for species number (S) to increase with sampled area (A) constitutes one of the most robust empirical laws of ecology, quantified by species-area relationships (SAR). In many ecosystems, SAR curves display a power-law dependence, $S \propto A^z$. The exponent z is always less than one but shows significant variation in different ecosystems. We study the multitype voter model as one of the simplest models able to reproduce SAR similar to those observed in real ecosystems in terms of basic ecological processes such as birth, dispersal and speciation. Within the model, the species-area exponent z depends on the dimensionless speciation rate ν , even though the detailed dependence is still matter of controversy. We present extensive numerical simulations in a broad range of speciation rates from $\nu = 10^{-3}$ down to $\nu = 10^{-11}$, where the model reproduces values of the exponent observed in nature. In particular, we show that the inverse of the species-area exponent linearly depends on the logarithm of ν . Further, we compare the model outcomes with field data collected from previous studies, for which we separate the effect of the speciation rate from that of the different species lifespans. We find a good linear relationship between inverse exponents and logarithm of species lifespans. However, the slope sets bounds on the speciation rates that can hardly be justified on evolutionary basis, suggesting that additional effects should be taken into account to consistently interpret the observed exponents.

Key words: Species-area exponents, Spatial models, Dispersal, Voter model, Biodiversity, Neutral Theory

1 Introduction

Species-area relationships (SAR) quantify ecosystem richness and, in particular, the spatial variations of biodiversity. These curves measure the average number of species (S) present in a sample area (A) of a given ecosystem and usually display a triphasic shape (Preston, 1960; Rosenzweig, 1995; Hubbell, 2001). For small areas (below the dispersal range) and large areas (continental scale), the number of species rapidly increases with the area; while for intermediate areas a slower, sub-linear growth is observed. The intermediate range is the most intriguing one and has gathered much attention since its discovery. Although many functional forms have been proposed to fit the data in this intermediate regime (He and Legendre, 1996; Tjorve, 2003), the most common and widely accepted ones are the algebraic law $S = CA^z$ (with $z < 1$ and C a positive constant) proposed by Arrhenius (1921) (see also Gleason (1922)), and the logarithmic one $S \approx C \ln A$ due to Fisher et al. (1943). A recent survey by Drakare et al. (2006), reconsidering most of the existing SAR studies from different ecosystems, shows that the former provides a better fit in about half of the cases. Even though any of the two hypothesis cannot be *a priori* discarded, much efforts across the years (Preston, 1962; MacArthur and Wilson, 1967; Connor and McCoy, 1979; Wright, 1988; Kohn and Walsh, 1994; Durrett and Levin, 1996; Hubbell, 2001; Chave et al., 2002; He and Legendre, 2002; Martin and Goldenfeld, 2006) have been devoted to explain the observed values of the exponent z . Observations support the idea of a dependence of the exponent z on quantities such as latitude (Allen and Gillooly, 2006) and body size of considered species (Drakare et al., 2006). Notwithstanding observations and theoretical efforts, a satisfactory theory able to predict the value of the exponent in different ecological situations is still lacking.

On the theoretical side, two distinct viewpoints on ecosystems organization correspond to different explanations for species-area relationships. According to the first, larger areas contain a larger variety of habitats and consequently can sustain a richer species diversity (Kohn and Walsh, 1994). For the second viewpoint, species-area relationships are the outcome of demographic processes such as colonization, dispersal, speciation and local extinction, and do not need environmental diversity for their explanation (MacArthur and Wilson, 1967; Hubbell, 2001). We should mention a third explanation, ascribing species-area relationships to statistical biases induced by the skewedness of species abundance distributions (He and Legendre, 2002; Martin and Goldenfeld, 2006). As both niche-based and neutral-dispersal mechanisms are able to sustain diversity, the hope is to extract information on the importance of the different classes of effects from the shape of SAR curves (Chave et al., 2002).

We consider the voter model as the simplest prototype of neutral models able

to generate non-trivial species-area relationships (Durrett and Levin, 1996; Hubbell, 2001; Zillio et al., 2005; Rosindell and Cornell, 2007). The model accounts, in a simple way, for the processes of birth, local dispersal and introduction of new species. Its main parameter, ν , is a dimensionless number measuring the rate of appearance of new species — speciation events — in units of the death rate. The other ingredient is the dispersal kernel, which quantifies the probability for an individual of a species to colonize different locations in the ecosystem. Speciation and dispersal are enough to produce triphasic SAR curves resembling those observed in field data (Chave et al., 2002; Rosindell and Cornell, 2007). In particular, for local (short-range) dispersal, the intermediate regime is well described by a power-law behavior $S = CA^z$, with an exponent z depending on the speciation rate ν . We mention that the logarithmic function is reproduced by the voter model with global dispersal, when individuals can invade all loci of the ecosystem (Coleman, 1981; Chave et al., 2002). The logarithmic law is also retrieved, for large dispersal, for areas being smaller than the dispersal range, indeed at these scales the dispersal appears as if it was long range. Understanding how the exponent z depends on the parameters of the model (in this case, mostly on ν) is fundamental to move a step toward the theoretical prediction of the variations of experimentally observed exponents in terms of ecological quantities.

However, this dependence has been source of some controversy in the literature. In a seminal paper, Durrett and Levin (1996) proposed a formula according to which, in the limit small ν , $z \sim 1/\ln(1/\nu)$. Rosindell and Cornell (2007) suggest a power law relationship between z and ν . Finally, the scaling argument of Zillio et al. (2005) predicts z approaching a finite value $z \approx 0.2$ for vanishing ν . These discrepancies have not yet been settled and, due to the weak dependence of z on ν , a clean answer requires numerical simulations with ν varying over several orders of magnitude. So far, only speciation rates $\nu \gtrsim 10^{-6}$ were explored as simulations at lower (possibly more realistic) values of the speciation rate are computationally very expensive.

In this paper we present results of simulations of the voter model with speciation rates varying in a wide range of values from $\nu = 10^{-3}$ down to $\nu = 10^{-11}$, with the twofold aim of disentangling the low speciation rate behavior and examine an ecologically relevant range of parameters. Our findings are also useful to assess whether neutral predictions are consistent with realistic speciation rates (Hubbell, 2001), a question which raised a heated debate (Hubbell, 2003; Ricklefs, 2003). SAR curves resulting from our simulations are characterized by a power law behavior with exponent z , displaying a logarithmic dependence on the speciation rate and supporting *de facto* Durrett-Levin’s scenario, even though with different numerical coefficients. In agreement with Rosindell and Cornell (2007), we also found that the exponent z is essentially insensitive to the dispersal range implying that, accepting the hypothesis of the model, the observation of a species-area exponent imposes strong con-

straints on the rate of appearance of new species. In the Discussion section, we examine the plausibility of the model predictions on the basis of data available in the literature. In particular, we consider the z -values reported in the literature for different taxa and, due to the absence of reliable data on speciation times, we study how measured exponents depend on the lifespan, with the additional assumption that average speciation times and lifespans are linked by a scaling relation. The observed variations turn out to be much larger than those allowed in the framework of the model; we finally discuss which effects may be included to possibly achieve a quantitative description.

2 Model

We consider the voter model with mutation as defined by Durrett and Levin (1996). Individuals belonging to different species are placed at each site of a two-dimensional ($L \times L$)-square lattice and evolve according to the following dynamics. At each time-step, a randomly chosen individual is killed, creating a gap which is immediately filled, with probability ν , by an individual from a new species (not present in the ecosystem) — speciation event — or, with a probability $1 - \nu$, by a new individual of an already existing (in the ecosystem) species chosen among those present in a neighborhood (that will be detailed below) of the site — birth/dispersal event. The dynamics is then advanced until the number of species in the ecosystem reaches a statistically steady value. Strictly speaking, the fact that empty locations are immediately colonized means that the birth rate is infinite (see discussion in Durrett and Levin (1996)). Therefore the basic time-step of the dynamics correspond to a death event, and thus the dimensionless parameter ν represents the speciation rate σ measured in unit of the death rate d . Equivalently, we can express ν as the average species lifespan t divided by the average time between speciation events $t^{(s)}$ (we shall come back to this point in the Discussion section).

As for the dispersal rule several options are possible. The simplest possibility is the nearest-neighbor rule, where the individual is replaced by one of the species present in the four neighbor sites with probability $1/4$. We will refer to this in the following as the nearest-neighbor (NN) case. A more realistic choice is to use a generic dispersal kernel introducing the probability $P(r)$ of a gap being filled by a species whose representative individual is at a distance r from it. We adopt a computationally simple instance by choosing the square kernel: we replace the individual with a copy of another individual randomly chosen in a square of side $2K + 1$ centered on the gap. This choice does not represent a restriction as it has been shown that the relevant quantity is the averaged square dispersal distance and not the specific functional form of the kernel. For instance, a Gaussian and a square dispersal kernel with the same squared dispersal distance produce very similar SAR (Rosindell and Cornell,

2007).

We stress that, independently of the dispersal rule, the model is completely neutral: all species (and individuals) undergo the same dynamics, as differences among species arise only due to demographic stochasticity.

Simulations have been efficiently performed by using the dual representation of the voter model (Holley and Liggett, 1975), providing a way to reconstruct the asymptotic configuration of the ecosystem by tracing backward in time its evolution. An important advantage of the dual representation is that it reconstructs the genealogy of each individual up to the speciation event originating its species, meaning that the system is ensured to have reached equilibrium. Moreover, it allows to implement open boundary conditions: the genealogy of an individual can be reconstructed also when its ancestors are outside the simulated area, which can thus be considered as a sample of a virtually infinite ecosystem Rosindell and Cornell (2007); Rosindell et al. (2008). This means that we can interpret ν as a *bona fide* speciation rate, since immigration from outside the system is included in the birth-dispersal process. However, long-range immigration events qualitatively different from local dispersal (i.e. seeds transported by birds) can be modeled as an higher “effective” speciation rate ν . We managed to optimize the algorithm to simulate the model for very low speciation rates, down to $\nu = 10^{-11}$. Details on how the simulations have been performed and the statistics have been collected can be found in Appendix A; see also Rosindell et al. (2008) for other possible improvements of the coalescence algorithm. As for the dispersal, we explored both the NN and the square kernels, for the latter K has been varied in the range $K = 3 - 64$, though we shall mostly present the results for $K = 7$ (see the discussion in the next section).

3 Numerical results

We begin studying SAR curves obtained at fixed dispersal range ($K = 7$) and varying ν , as shown in Fig. 1. All curves display a fast growth for small areas with a crossover, for areas of the order of the dispersal kernel ($A \approx K^2$), to the power-law regime. The final regime where the number of species becomes linear with the area can be detected only for rather large speciation-rate values, $10^{-5} \leq \nu \leq 10^{-3}$; to observe it at lower values of ν much larger simulation samples would be required. In the inset, we plot the “local species-area exponent” for each curve, $d(\ln S)/d(\ln A)$, which clearly shows that the smaller ν the smaller the exponent becomes and the larger is the range of scales where a well defined power-law behavior establishes. Finally, when the parameter ν is not too small, it is possible to observe also the final linear regime which occurs for areas much larger than ν^{-1} (Durrett and Levin, 1996).

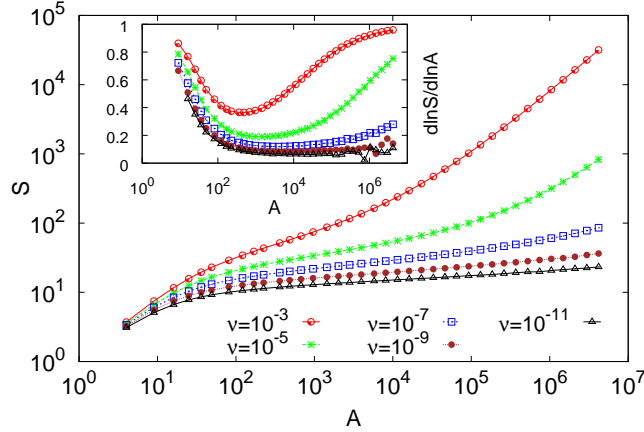


Fig. 1. Dependence of SAR curves on the speciation rate: S vs A for different speciation rates $\nu \in [10^{-9} : 10^{-3}]$ obtained with the square kernel with $K = 7$ in a simulation sample of side $L = 2048$, and averaging over 100 independent realizations. Note the triphasic shape clearly observable for larger values of ν . Inset: Logarithmic derivatives of the SAR curves, $d(\ln S)/d(\ln A)$ vs A , a plateau identifies the intermediate regime and the plateau value the exponent z . Note that z increases with ν and the intermediate regime enlarges in width at decreasing ν up to invading almost all the simulation sample for small ν values.

Figure 2 (left and middle panel) exemplifies the behavior of species-area curves at fixed ν and different dispersal range K . At increasing the dispersal range the onset of the power law regime shifts at larger areas, apparently without affecting the exponent. A more careful analysis of the local exponents $d(\ln S)/d(\ln A)$, shown in the right panel, detects a dependence of the value of the exponent on the dispersal range when this is small, $K \lesssim 5$, including the NN case.

On the other hand, when $K \gtrsim 5$, we did not observe any appreciable corrections to the value of the exponent. The independence against variations of K , when it is large enough, has been quantified by Rosindell and Cornell (2007), who have shown that curves obtained with different (not too small) K can be rescaled on a universal function of A and ν only via the transformation:

$$S = f(A, \nu, L) = K^r \phi(A/K^r, \nu) \quad (1)$$

characterized by the scaling exponent $r \approx 1.97$. We checked that this relation holds also with the small values of ν that we studied, for instance the insets of Figure 2 (left and middle panel) show it for $\nu = 10^{-5}$ and $\nu = 10^{-8}$. We will then study in the following the NN and the $K = 7$ cases, the former being that originally studied by Durrett and Levin (1996) and the latter being representative of the behavior of the model for large average dispersal distances.

We now turn to the main results of this paper about the dependence of z

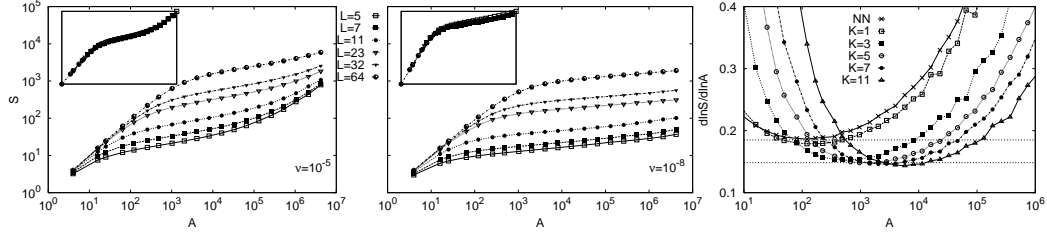


Fig. 2. Dependence of SAR curves on the dispersal length. Left and Middle panels: S vs A , for $\nu = 10^{-5}$ and $\nu = 10^{-8}$ obtained with the square kernel with several values of K in a simulation sample with $L = 2048$. Note that in the intermediate regime, which starts and ends at different area values by changing K , the slope of the curves is the same indicating that z is independent of K . Insets show the curves rescaled with the transformation (1). Right panel: local slopes $d(\ln S)/d(\ln A)$ vs A for $\nu = 10^{-6}$ and different kernels: NN and square with $K = 1, 3, 5, 7, 11$. Note the tiny dependence of the exponent (the region between the two horizontal lines) on the dispersal length. For $K \gtrsim 5$ the plateau region (lower horizontal line) does not change anymore apart from an horizontal shifting of its onset towards larger areas.

on ν . In Figure 3 (left) we show the exponent z as a function of the speciation rate ν (see Appendix A for a discussion on how we estimated z). We observe a clear discrepancy with previous predictions (Durrett and Levin, 1996; Rosindell and Cornell, 2007) (also shown in the picture). In particular, for $\nu \ll 1$, we found the data to fall into a straight line when plotting $1/z$ vs. $\ln(\nu)$ (Fig. 3 right), suggesting the following functional dependence

$$z = \frac{1}{q + m \ln(\nu)}, \quad (2)$$

by which we obtained a best fit to the data with $q \approx -3.3$ and $m \approx -0.72$. To compare our results with previous studies of these models, notice that the power-law fit suggested in by Rosindell and Cornell (2007) agrees with the data in the same range of speciation rate values, i.e. $\nu \geq 10^{-5}$. Deviation from a power law behavior are clearly observed for lower values of ν , where the data also rule out the saturation at $z \approx 0.2$ predicted by Zillio et al. (2005). Actually, our fit confirms Durrett and Levin (1996) prediction of a logarithmic decay of z with ν , up to corrections order $\mathcal{O}(\ln(\ln(\nu)))$. However, the fitting parameters m and q for both the square kernel with $K = 7$ and the NN kernel are very different from those of Durrett and Levin (see caption of Fig. 3). We conjecture that the differences in prefactors could be caused by two different assumptions used by Durrett and Levin to derive the dependence of z on ν . The first is about pre-asymptotic effects: the statistical results used by Durrett and Levin are strictly valid only when $t \rightarrow \infty$ which requires $\nu \rightarrow 0$, while finite-time corrections may affect the exponent value. In this respect, also for our data the $\nu \rightarrow 0$ limit seems to be crucial for the validity of the fit (2). The second is the assumption that a power law regime establishes from $A = 1$ to $A = \nu^{-1}$. Conversely, we observe the onset of the power law for areas being slightly larger than 1 even in the NN case. Moreover, the crossover to

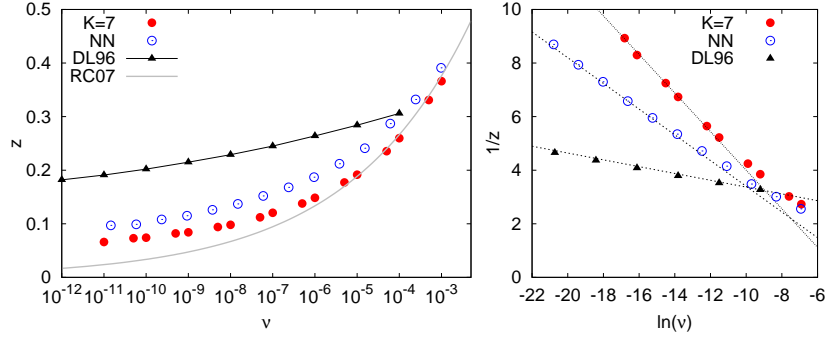


Fig. 3. Dependence of the species-area exponent on the speciation rate. (left) SAR exponent z versus the speciation rate ν for the NN and square kernel with $K = 7$. For comparison it is also shown the prediction by Durrett and Levin (1996) and the power-law dependence suggested by Rosindell and Cornell (2007). (right) Same data as left but shown plotting $1/z$ vs $\ln(\nu)$. The straight lines reports the best fit with (2) with parameters: $m = -0.48 \pm 0.02$, $q = -1.4$ for NN data, $m = -0.72 \pm 0.02$, $q = -3.2$ for square kernel data ($K = 7$) and $m = -0.127 \pm 0.002$, $q = 2.10$ for Durrett-Levin tabulated data. Errors on the exponents are of the order of symbol sizes, see Appendix A for details.

the linear asymptotic regime begins for areas quite smaller than $1/\nu$.

It should be noticed that discrepancies in the numerical factors have profound implications when the model is used to estimate a speciation rate from an observed species area exponent. The logarithmic dependence of z on ν makes, in fact, ν exponentially dependent on z . We will discuss in the next section how this dependence can be compared with experimental data. It should also be remarked that both Durrett and Levin prediction and equation (2) are valid for small values of ν and can lead to incorrect results, such as negative z , for ν close to 1.

All simulations so far presented have been performed with open boundary conditions, which are appropriate when the sample is a homogeneous portion of a much larger ecosystem. However, closed boundary conditions can be of interest for modeling confined ecosystems such as islands. Intuitively, open boundaries allow new species to immigrate into the sampled system from the external infinite ecosystem, independently of the speciation events. Closed boundaries exclude this possibility and are thus expected to reduce the exponent z and, in general, species richness. Fixing the speciation rate ν the decreasing of z becomes more and more efficient as the system size decreases, and at fixed size the effect is the stronger the smaller is ν . The closed boundaries effects becomes dramatic for simulation samples with $A \approx 1/\nu$, and, in particular, we observed that for ν such that $A \lesssim 1/\nu$, biodiversity is definitely lost, i.e. $z = 0$ (see Fig 4). Notice that when the system size is large (i.e. $A \geq 2048^2$) and the speciation rate is not too small (i.e. $\nu \leq 10^{-4}$) the exponent is essentially the same of the open boundary case. We also remind that, if islands are modeled

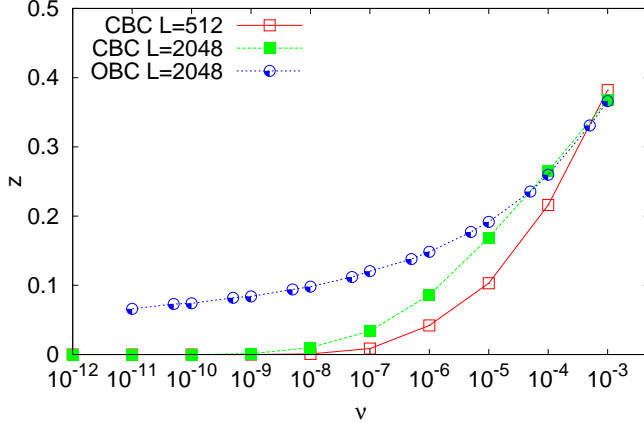


Fig. 4. z vs ν for the square kernel with $K = 7$ with closed boundary conditions (CBC) at $L = 512$ and $L = 2048$ compared with the result with open boundary conditions (OBC) and $L = 2048$. Note how the decrease of z with ν becomes more and more dramatic at decreasing the size.

by closed boundaries, the parameter ν should be meant to include the immigration rate of new species (MacArthur and Wilson, 1967), since dispersal from outside the system is forbidden in this case.

We stress here that the implementation of closed boundary is a very simplistic way of describing confined ecosystems and more sophisticated effects may be relevant in these cases. For example, it has been shown that the dynamics at the edge of ecosystems can be quite different from that in the bulk (Laurance et al., 2006). This means that “open” situations, when the sample is part of a much larger and homogeneous system, provide a much safer comparison between models and field data.

4 Comparison with empirical data and discussion

Species-Area relationships have been subject of intense experimental research in a variety of ecosystems, and the range of variability of the exponent describing the intermediate regime goes from $z \approx 0.05$ in bacteria (Horner-Devine et al., 2004) to $z = 0.4 - 0.5$ in some plants community (see Drakare et al. (2006) for an exhaustive review of field observations). According to our results, the voter model with speciation displays a variability of z in the same range when the speciation rate is allowed to vary over several orders of magnitude. It is thus tempting to go in the direction of a more quantitative comparison between field data and the model results.

As a first step, we assume that the model is able to describe the main features of groups of trophically similar species and explore the consequences of

this assumption. This requires that a speciation rate and a dispersal range for the whole group of species can be properly defined, although we know that speciation rates (Mariakeva and Gorshkov, 2004) and dispersal ranges (Nathan and Muller-Landau, 2000; Kinlan and Gaines, 2003) may have significant variations from species to species. However, in the model the exponent is essentially independent on the dispersal range and depends only logarithmically from the speciation rate, so that these variations might be disregarded treating all species in the group as having the same “average” dispersal range and speciation rate.

As far as dispersal is concerned, we only found a tiny dependence for very short dispersal range, around $K < 5$. Above these values the exponent is independent of the dispersal range confirming previous results Rosindell and Cornell (2007). Realistic average dispersal ranges (Nathan and Muller-Landau, 2000; Kinlan and Gaines, 2003) are certainly far from the short dispersal case, due to animal motility or wind for seeds. Therefore, we assume that the dispersal range of real groups of species is always in the range where the exponent is dispersal-independent. It is however worth remarking that the dispersal range can still affect the spatial biodiversity via the power-law prefactor, whose increase can lead to a large number of species that, when z is small, increases very slowly with the area. In this respect, the model outcomes are in contrast with interpretations of low values for z in bacteria as an effect of large dispersal distances as argued in Drakare et al. (2006) and Horner-Devine et al. (2004).

What about speciation? Unfortunately, we do not have ecological data allowing us to directly estimate the frequency of speciation events. Data from fossils suggest an average speciation rate on Earth of about three species per year (Sepkoski, 1998), but it is hard to infer from this number a reasonable rate for a living system. Also estimates based on mutation rates (Mariakeva and Gorshkov, 2004) could be flawed due to genetic bottlenecks and phenomena like horizontal gene transfer (Jain et al., 1999). Moreover, as discussed in the Model section, the parameter ν should be interpreted as an “effective” speciation rate, incorporating also long-range dispersal events. Within the model framework, our results show that species-area exponent and dimensionless speciation rate ν are related even when the latter is very small, implying that an observed value of the exponent z would predict the rate of introduction of new species ν . Remarkably, the existence of positive correlations of these two quantities is consistent with observational results. As an example, it is known that close to the equator species-area exponents tend to increase (Drakare et al., 2006) together with speciation rates (Allen and Gillooly, 2006) and overall biodiversity (Stevens, 1986).

In order to test the ecological plausibility of the relation between z and ν , we make use of the definition of ν as the ratio between the speciation rate σ

and the death rate d . From Eq. (2) and separating the contribution from the variation in the speciation rate from that of the variation in the death rate, we have

$$\frac{1}{z} = q + m \ln(\nu) = q + m[\ln(\sigma) - \ln(d)], \quad (3)$$

where the arguments of the logarithm are made dimensionless by measuring them in the same units. To ease the interpretation, we recast this equality in the time domain using the lifespans $t = 1/d$ and the average time between speciation events $t^{(s)} = 1/\sigma$:

$$\frac{1}{z} = q + m[\ln(t) - \ln(t^{(s)})], \quad (4)$$

The first term on the right hand side accounts for the variation in z due to the lifespan which is, of course, much easier to estimate than the term due to speciation time and can still be important and informative. Indeed there are evidences that taxa having a shorter generation time have generally lower species area exponents (Horner-Devine et al., 2004; Green et al., 2004; Zhou et al., 2008) (we recall that m is negative). We thus study how the inverse exponent $1/z$ varies with the logarithm of the lifespan. The results of this analysis are presented in figure 5 for data obtained from the literature (see Appendix B for a description of how data have been collected), which shows that a linear relationship fits rather well the data, with an observed slope $m_{meas} = -1.76 \pm 0.13$ (dashed line in the figure) which is different from $m \approx -0.72$ predicted by the voter model.

The fact that for species-area exponents measured in field data we found $1/z \propto \ln(t)$ suggests a scaling relationship between speciation time and lifespan, i.e. that $t^{(s)} \sim t^\gamma$, so that $m_{meas} = m(1 - \gamma)$, as clear by substitution in the previous formula. We do not have any *a priori* explanation for justifying a power law dependence of the speciation time on the lifespan, apart from the observation that the variations of many ecologically relevant rates among species are governed by scaling laws (Brown et al., 2004). We are not aware of independent estimation of the dependence of the speciation time on the lifespan so to confirm or reject the outcome of our analysis.

However, the relation $m_{meas} = m(1 - \gamma)$, with m and m_{meas} fixed by the voter model and field data respectively, yields a negative γ . This result is in contrast with biological expectations as it would imply, e.g., a speciation time for bacteria much longer than the one for trees, which is hard to justify biologically. Reasonable expectations would have been $0 < \gamma < 1$. The limiting case of $\gamma = 1$ is the trivial case in which speciation time is proportional to the lifespan. This would have lead to $m_{meas} = 0$, i.e. same $\nu = t/t^{(s)}$ and z for all taxa. The other limiting case is $\gamma = 0$, which is plausible when the possibility of creating a new species is triggered by some external mechanism, like the availability of new niches, which is not strongly correlated to any particular

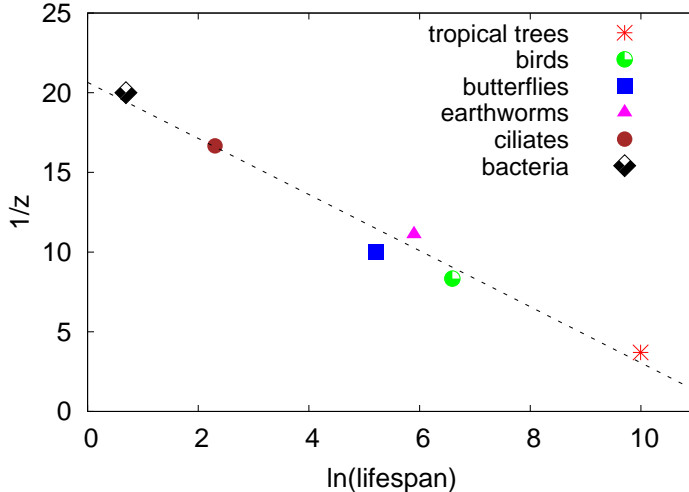


Fig. 5. $1/z$ vs logarithm of lifespans (measured in days) for several taxa. Dashed line is the best fit, giving a slope $m_{meas} = -1.76 \pm 0.13$ (reduced $\chi^2 = 1.04$).

feature of the species. Another justification could come from co-evolutionary mechanisms: species having very different lifespans can still evolve on similar timescales due to their ecological interactions (Thompson and Jeremy, 1992). Actually, co-speciation is known to occur in some cases of host-parasite systems (Clayton et al., 2003). In this case, one would find $m_{meas} = m$.

The inconsistent value of γ we obtained can be interpreted either as a failure of the basic assumptions of the neutral model and thus of its inadequacy in describing realistic ecosystems or as the need of additional mechanisms to be included in the framework of dispersal models. The quality of the linear fit shown in Fig. 5 would suggest to opt for the second interpretation, even if the linear dependence of $1/z$ on $\log(t)$ needs to be tested with further measures. In recent years, several attempts of relaxing the strong assumptions of neutral models have been tried. The results of these models are pretty robust with respect to modifications of some hypothesis such as the saturation of the resources (Etienne et al., 2007b). In spatial models, it has been also observed that the introduction of trade-offs does not have a dramatic effect on species-area exponents (Chave et al., 2002). Therefore, it is reasonable to search for other elements in the model which can lead to a failure in reproducing the observed data. In particular, the assumptions of a point speciation mode (i.e. the fact that each individual has a fixed probability to speciate) is known to be crucial and the results may change dramatically when considering “fission modes”, corresponding to allopatric speciation (Hubbell, 2001, 2003; Ricklefs, 2003; Etienne et al., 2007a). Indeed, it seems like the most important assumption of neutral models is that new species enter the system with a population of a single individual (Zillio and Condit, 2007). This could explain why speciation rates predicted by neutral models with point speciation may look unrealistically high: new species are introduced with only one individual

and they have an high probability of going extinct before being able to grow. In other words, there could be a discrepancy between the parameter ν in the model and the experimentally observed speciation rate. The effect of realistic speciation mechanisms on neutral and more general dispersal models could be key to understand the puzzle of the observed variation of the exponents among different taxa.

In conclusion, simulations of the multi-type voter model for low values of the speciation rate show a clear logarithmic functional dependence of the specie-area exponent on the speciation rate and independence on the dispersal kernel (provided it is not too short ranged). Analysis of field data support a logarithmic dependence of the exponent z on the timescales of the problem, though with a prefactor which is incompatible with that found in the model. Our analysis points out that more refined models should allow larger variations in the exponent z in order to be consistent with observational data.

5 Acknowledgments

We would like to thank K. H. Andersen and F. Cecconi for a critical reading of the manuscript.

A Simulations and data analysis

By means of the dual representation of the voter model, the model becomes equivalent to a system of coalescing random walkers, where a birth/dispersal event corresponds to collision-coalescence of two walkers and a speciation event to the death of a walker. Therefore, as time proceeds, less and less surviving walkers should be accounted, speeding up the simulation. This allowed us to simulate lattices $L \times L$ with $L = 1024, 2048$ and $L = 4096$ for $\nu \in [10^{-11} : 10^{-3}]$. In order to test the effect of the system size when boundaries are present, we have also used $L = 512$; in this case we generalized the algorithm by just refusing all moves causing the exit of a walker from the simulation domain, thus constraining the walkers to remain inside the initial grid. To embank the simulation bottleneck due to the initial presence of L^2 walkers we optimized the walker collision detection by means of a look-up table. For each ν and L we repeated the simulation many times with different seeds for the random number generator, typically from 100 to 150 – 300 for $L = 2048, 4096$ and $L = 1024, 512$, respectively. For the lower values of ν simulations get very slow and a lower number of realization was used, typically from 20 to 60 for $\nu < 10^{-10}$. Once the species occupancy patterns are obtained SAR curves are derived by averaging the number of species in non-overlapping squares of

side $\sqrt{A} = 1, \dots, L$ whose union completely covers the simulation grid. So that averages are performed both over the number of sampled areas in each realization and over different realizations. Statistical errors on the average are also computed. The exponent z characterizing the power law growth of S with A is then estimated fitting by a linear least square method the function $\ln S = q + z \ln A$ with $A \in [A_{min} : A_{max}]$ chosen at the beginning and the end of the intermediate regime, respectively. The fit were performed by minimizing the reduced χ^2 function (i.e. normalizing the χ^2 with the number of degrees of freedom) but constraining the minimal number of points to be considered (from 5 to 15 depending on the extension of the intermediate range). As the least square error is smaller than the variability of the fitted z at changing the minimal number of constrained points, we set the error on the estimate as such variability. In Fig. 3 errors are comparable with the symbol size.

The quality of the fit is then compared by a direct inspection of the local slopes (logarithmic derivatives, i.e. $d \ln S / d \ln A$ vs $\ln A$) of the SAR curves.

B Details on observational data

Data presented in Fig. 5 are based on the collection of exponents presented by Horner-Devine et al. (2004). We avoided presenting the z -value for plants since it varies a lot among different studies (see supplementary information of Horner-Devine et al. (2004)). As a representative of long-lived organisms we have chosen tropical forest trees which are well studied and we assumed for them $z = 0.27$ (Lonsdale, 1999) and average lifespan ≈ 60 years (Condit et al., 1999). The exponent z values for butterflies, earthworms, birds and ciliates are the same of the original reference; the value for bacteria $z = 0.05$ is an average between the value $z = 0.04$ in Horner-Devine et al. (2004) and the value $z = 0.06$ in a more recent study by Zhou et al. (2008). As far as the other lifespans are concerned, we must stress that it can vary much from species to species and, in most cases, it is hard to find in the literature good statistical studies. However, due to the logarithmic dependence, the fit is not be so sensitive to errors in the estimates as far as the order of magnitude is correct. We assumed an average lifespan of 2 years for birds (see, e.g., Speakman (2005)). Despite their short average lifespan in their adult stage, butterflies usually have a few generations per year (Gilbert and Singer, 1975); we assumed an average of two, corresponding to $t = 0.5$ years. Other estimated lifespans are 1 year for earthworms (Edwards and Bohlen, 1977), 10 days for ciliates (Jensen and Verschoor, 2004) and 2 days for bacteria (Clarholm and Rosswall, 1980).

References

- Allen, A. P., Gillooly, J. F., 2006. Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. *Ecol. Lett.* 9, 947-954.
- Arrhenius, O., 1921. Species and area. *J. Ecol.* 9, 95-99.
- Brown, J.H., et al. 2004. Towards a metabolic theory of ecology. *Ecology* 85, 1771-1789.
- Chave, J., Muller-Landau, H. C., Levin, S. A., 2002. Comparing Classical Community Models: Theoretical Consequences for Patterns of Diversity. *Am. Nat.* 159, 1-23.
- Clarholm, M., Rosswall, T., 1980. Biomass and turnover of bacteria in a forest soil and a peat. *Soil Biol. Bioch.* 12, 49-57.
- Clayton, D. H., Bush, S. E., Goates, B. M., Johnson, K. P., 2003. Host defense reinforces host-parasite cospeciation. *Proc. Natl. Acad. Sci. USA* 100, 15694-15699.
- Coleman, B. D. 1981. Random Placement and Species-Area Relations. *Math. Biosci.* 54, 191-225.
- Condit, R., et al. 1999. Dynamics of the forest communities at Pasoh and Barro Colorado: comparing two 50-ha plots. *Phil. Trans. R. Soc. London B* 354, 1739-1748.
- Connor, E. F., McCoy, E. D., 1979. The statistics and biology of the Species-Area Relationship. *Am. Nat.* 113, 791-833.
- Drakare, S., Lennon, J. J., Hildebrandt, H., 2006. The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecol. Lett.* 9, 215-227.
- Durrett, R., Levin, S. A., 1996. Spatial Models for Species-Area Curves. *J. Theor. Biol.* 179, 119-127.
- Edwards, C. E., Bohlen, P. J., 1977. *Biology and Ecology of Earthworms*. Chapman and Hall, London.
- Etienne, R. S., Apol, M. E. F., Off, H., Weissing, F. J., 2007a. Modes of speciation and the neutral theory of biodiversity. *Oikos* 116, 241-258.
- Etienne, R. S., Alonso, D., McKane, A. J., 2007b. The zero-sum assumption in neutral biodiversity theory. *J. Theor. Biol.* 248, 522-536.
- Fisher, R. A., Corbet, A. S., Williams, C.-B., 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. An. Ecol.* 12, 42-58.
- Gilbert, E., Singer, M. C., 1975. Butterfly ecology. *Annu. Rev. Ecol. System.* 6, 365-395.
- Gleason, H. A., 1922. On the relation between species and area. *Ecology* 3, 158-162.
- Green, J. L., et al. 2004 Spatial scaling of microbial eukaryote diversity. *Nature* 432, 747-750.
- He, F., Legendre, P., 1996. On Species-Area Relations. *Am. Nat.* 148, 719-737
- He, F., Legendre, P., 2002. Species diversity patterns derived from species area models. *Ecology* 83, 1185-1198.

- Holley, R., Liggett, T. M., 1975. Ergodic theorems for weakly interacting systems and the voter model. *Ann. Prob.* 3, 643-663.
- Horner-Devine, M. C., Lage M., Hughes, J. B., Bohannon, B. J. M., 2004. A taxa-area relationship for bacteria. *Nature* 432, 750-753.
- Hubbell, S. P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.
- Hubbell, S. P., 2003. Modes of speciation and the lifespans of species under neutrality: a response to the comment of Robert E. Ricklefs. *Oikos* 100, 193-199.
- Jain, R., Rivera, M. C., Lake, J. A., 1999. Horizontal gene transfer among genomes: The complexity hypothesis. *Proc. Natl. Acad. Sci. USA* 96, 3801-3806.
- Jensen, T. C., Verschoor, A. M., 2004. Effects of food quality on life history of the rotifer *Brachionus calyciflorus* Pallas. *Freshwater Biology* 49, 1138-1151.
- Kinlan, B. P., Gaines, S. D., 2003. Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology* 84, 2007-2020.
- Kohn, D. D., Walsh, D. M., 1994. Plant species richness - The effect of island size and habitat diversity. *J. Ecol.* 82, 367-377.
- Laurance, W. F. et al., 2006 Rapid decay of tree-community composition in Amazonian forest fragments. *Proc. Natl. Acad. Sci. USA* 103, 19010-19014.
- Lonsdale, W. M., 1999. Global Patterns of Plant Invasions and the Concept of Invasibility. *Ecology* 80, 1522-1536.
- MacArthur, R. H., Wilson, E. O., 1967. *The Theory of Island Biogeography*. Princeton University Press, NJ.
- Mariakeva, A. M., Gorshkov, V. G., 2004. On the dependence of speciation rates on species abundance and characteristic population size. *J. Biosci.* 29, 119-128.
- Martin, H. G., Goldenfeld, N., 2006. On the origin and robustness of power-law species-area relationships in ecology. *Proc. Natl. Acad. Sci. USA* 103, 10310-10315.
- Nathan, R., Muller-Landau, H.C., 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends Ecol. Evol.* 15, 278-285.
- Preston, F. W., 1960 Time and space and the variation of species. *Ecology* 41, 612-627.
- Preston, F. W., 1962 The canonical distribution of commonness and rarity. *Ecology* 43, 185-215, 410-432.
- Ricklefs, R. E., 2003. A comment on Hubbell's zero-sum ecological drift model. *Oikos* 100, 185-192.
- Rosenzweig M. L. 1995. *Species Diversity in Space and Time*. Cambridge University Press, Cambridge.
- Rosindell, J., Cornell, S. J., 2007. Species-area relationships from a spatially explicit neutral model in an infinite landscape. *Ecol. Lett.* 10, 586-595.
- Rosindell, J., Wong, Y., Etienne, R. S., 2008. A coalescence approach to spatial neutral ecology. *Ecol. Informatics* 3, 259-271.

- Sepkoski, J. J., 1998. Rates of speciation in the fossil record. *Phil. Trans. R. Soc. London B* 353, 315-326.
- Speakman, J. R., 2005. Body size, energy metabolism and lifespan. *J. Exp. Biol.* 208, 1717-1730.
- Stevens, G. C., 1986. The latitudinal gradient in geographical range: how so many species coexist in the tropics. *Am. Nat.* 133, 240-256.
- Thompson, J. N., Jeremy, J., 1992. Gene-for-gene coevolution between plants and parasites. *Nature* 360, 121-125.
- Tjorve, E., 2003. Shapes and functions of species-area curves: a review of possible models. *J. Biogeog.* 30, 827-835.
- Wright, S. J., 1988. Patterns of abundance and the form of the species-area relation. *Am. Nat.* 131, 401-411.
- Zhou, J., Kang, S., Schadt, C. W., Garten, C. T., 2008. Spatial scaling of functional gene diversity across various microbial taxa. *Proc. Natl. Acad. Sci. USA* 105, 7768-7773.
- Zillio, T., Volkov, I., Banavar, J. R., Hubbell, S. P., Maritan, A., 2005. Spatial Scaling in Model Plant Community. *Phys. Rev. Lett.* 95, 098101.
- Zillio, T., Condit, R., 2007. The impact of neutrality, niche differentiation and species input on diversity and abundance distributions. *Oikos* 116, 931-940.